GS01 0163
Analysis of Microarray Data

Keith Baggerly and Brad Broom
Department of Bioinformatics and Computational Biology
UT M. D. Anderson Cancer Center
kabagg@mdanderson.org
bmbroom@mdanderson.org

1 September 2009

Lecture 0: Outline and Details

Welcome!

- Where are we?

- Why are we here?

- Who are we?

- What are we going to tell you about, and when?

- What do we assume you're familiar with?

- Where and when can you find ???

- What are we going to ask of you?

So, where are we?

We were initially scheduled to meet in BSRB S3.8367 (the "small classroom"), but we have been moved to the library, BSRB S3.8351.

We will meet in the small classroom on three days: Tues, September 15, Thurs Oct 1, and Thurs Dec 3.

Course meets Tuesdays and Thursdays from 3-4:30, starting September 1 and running through the middle of December (15 weeks, with one scheduled off day for Thanksgiving).

So, why are we here?

We want to learn about microarrays.

Microarrays have caught on as high-throughput assays for understanding molecular biology – they let us measure expression levels for thousands of genes in a single sample all at once.

This means new biology, and new data analysis.

We want to

- Understand how microarrays work and how they are analyzed.

- Perform some basic analyses of microarrays.

We're going to dive in here; we hope to have you analyzing data this week or next.

Who are we?

Keith Baggerly, FCT 4.5016, kabagg@mdanderson.org, (713) 563-4290

Brad Broom, FCT 4.5006, bmbroom@mdanderson.org (713) 792-2617

We're faculty in the department of Bioinformatics and Computational Biology at MD Anderson. We've been working with microarray data for about 10 years now. We think we've done pretty well with it, and we're going to trade off in telling you about it.

What are we going to tell you about, and when?

Week 1: How microarrays work, and Affymetrix file structures. The basics of dChip.

Week 2: Using dChip. Connecting numbers to biology – the basics of annotation and databases. Documentation.

Week 3: Reviewing initial analyses. Introduction to R.

Week 4: Sample analysis using R/Bioconductor for Affymetrix.

Week 5: Reviewing R basics, quantification and normalization of Affymetrix data, Latin Square studies.

Week 6: Measures of differential expression (DE); R and DE.

What are we going to tell you about, and when?

Week 7: Experimental Design, linear models

Week 8: Clustering microarray data, diagnostics, validation.

Week 9: Classification.

Week 10: Genome Browsing, Pathways

Weeks 11-12: GSEA, Pooling Data

Week 13: Other Assays

Week 14: Reading the Literature (Thanksgiving)

Week 15: Student Presentations, TCGA

What do we assume you're familiar with?

Some biology.

Some programming.

Some statistics.

Computing

Microarray datasets are large. One of our initial data sets will involve 100 Affy gene chips; the files take up 1.5G of disk space when uncompressed. Another dataset takes 4G.

The image files from some of the newest arrays (e.g., Affymetrix exon arrays) are 1.5G per array, so data summarization is a real issue.

We will be using lots of freeware, including dChip (Windows-based), R, and the Bioconductor repository.

We will use version 2.9.1 of R, and version 2.4 of BioConductor.

Textbooks

Optional: Gentleman R, et al (eds). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* Springer-Verlag, New York, 2005.

Optional: Simon RM, et al. *Design and Analysis of DNA Microarray Investigations.* Springer-Verlag, New York, 2003.

Optional: Dalgaard P. *Introductory Statistics with R.* Springer-Verlag, New York, 2002.

Optional: Speed T (ed). *Statistical Analysis of Gene Expression Microarray Data.* Chapman and Hall, New York, 2003.

Optional: Parmigiani G, et al (eds). *The Analysis of Gene Expression Data.* Springer-Verlag, New York, 2003.

Where can you find ???

- Course web site `http://bioinformatics.mdanderson.org/MicroarrayCourse`

- Office hours: T, Th, 4:30-5:30, HMB
  - Dr. Baggerly: FCT 4.5016
  - Dr. Broom: FCT 4.5006

What are we going to ask of you?

Grading: Homeworks (roughly every two weeks)

1. Homeworks can be worked on jointly.

2. Assignments submitted electronically, in a single file (zipped is fine), by midnight on the assigned due date.

3. Submissions should include figures, R code, and text.

We're very interesting in getting good documentation for how analyses were performed. Thus, as we move into R we'll also be talking about *Sweave*, which helps us track things better.

A project.

An exam.